DOCUMENT RESUME

ED 179 859 CG 014 000

AUTHOR Halpin, Gerald: And Others
TITLE Effects of Study and Testing on Student Retention.
PUB DATE Sep 79
NOTE 14p.: Paper presented at the Annual Convention of the American Psychological Association (87th, New York, NY, September 1-5, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement: Behavior Patterns: College Students: Comparative Analysis: *Learning Processes: Research Projects: *Retention: Retention Studies: *Student Behavior: Study Habits: *Testing: *Test Wiseness

ABSTRACT

The effect of testing on student retention was investigated to determine if it was the actual test taking, the prior preparation for the test, or the combination of studying for and taking the test which actually affected retention. Subjects were undergraduate students (N=90) who were divided into two different study conditions, i.e., test and no-test. The test condition consisted of two different kinds of test, multiple-choice and short-answer, each containing two different levels of item complexity, knowledge and concept. Results indicated that: (1) subjects in the study condition test group scored higher on retention than subjects in the study condition no-test group: (2) study condition test subjects who took either the multiple-choice or short-answer test scored higher than subjects who took no test: (3) means for the multiple-choice test and short-answer test groups did not significantly differ: and (4) the short-answer test treatment group failed to differ from the no-test treatment group. (Author/HLM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED179859

EFFECTS OF STUDY AND TESTING ON STUDENT RETENTION

Gerald Halpin, Glennelle Halpin, and Edith Miller

Auburn University

Summary of research presented in poster session
at the meeting of the American Psychological Association,
New York, September 1979.

CG 014000

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gerald Halpin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Summary

In 1885 Ebbinghaus published the first experimental study of human memory. There have been numerous studies in the fields of education and psychology since that time which have addressed the problems of learning and retention. This research generally falls into two basic categories: (a) studies concerned with theory which focus on the recall of word pairs, series of words, sentences, and the like and (b) more practical research which includes studies of educational achievement and the retention of learning. A survey of the literature revealed far more studies in the former category than in the latter. An even greater deficit was found in the number of studies of specific educational practices related to retention of classroom learning. Harnquist (1977) highlighted this deficiency in an invited address to the 1977 meeting of the American Educational Research Association which he called "Enduring Effects of School--A Neglected Area in Educational Research."

Our study is one in this neglected area. It was designed to investigate the effect of a particular classroom activity--testing--on student retention. More specifically, the purpose of this study was to determine if it is the actual taking of the test, the prior preparation for the test, or a combination of studying for and taking the test which affects student retention if such an effect exists. Further purposes of this study were to determine if type of criterion measure (multiple-choice or short answer) or complexity of test item (knowledge or concept) significantly moderates the effects of testing on retention.

Method

Subjects

Subjects for this study were 90 undergraduate students enrolled in five educational psychology classes at a large southern university which

attracts students with diverse backgrounds from an extensive geographical area. Included were both female (N = 60) and male (N = 30) education (N = 62) and noneducation (N = 28) majors (median age = 20 years).

Procedure

Treatment. Treatment in this experiment consisted in part of two different study conditions: test and no test. In the test condition, subjects read the text assignment and attended class with the expectation of being tested. In the no test condition, subjects were asked to "read the text assignment and attend class to learn rather than for a test."

Treatment in this study further consisted of two different kinds of tests, multiple-choice (30 items) and short answer (30 items created from the multiple-choice item stems), each containing two different levels of item complexity, knowledge (15 items) and concept (15 items).

Experimental Procedures. Using class rolls, subjects within each class were randomly assigned to one of six groups and each group was randomly assigned (a) study condition and (b) test treatments as follows:

Group 1: (a) test, (b) multiple-choice

Group 2: (a) test, (b) short answer

Group 3: (a) test, (b) no test

Group 4: (a) no test, (b) multiple-choice

Group 5: (a) no test, (b) short answer

Group 6: (a) no test, (b) no test

During the introductory meeting of each of the educational psychology classes students were given by their instructor a syllabus with assignments and test dates (study condition-test). Each instructor subsequently called the names of those students who had been assigned to study condition-no test

and asked them to stay briefly after class where they met with an experimenter who explained that they had been selected to participate in an evaluation of ongoing instructional methods. For their participation, which they were not to discuss with anyone, they would be given an "A" in lieu of their earned grade on the upcoming test. They were, however, still asked to "read the textbook assignment and attend class but in order to learn rather than for a test."

Each instructor (white females) taught behavioristic learning theory to her respective class using a lecture-discussion approach for the next two weeks. On the assigned test day two of the experimenters came to the classroom and asked for by name those students who had been assigned to the study condition-no test group as well as those in Group 3--study condition-test, test condition-no test.

Subjects in Group 3 and Group 6--study condition-no test, test condition-no test went with one experimenter to a vacant classroom where they were told, as some of the group already knew, that their class was participating in an evaluation of instructional techniques which would necessitate their not responding to the test their classmates were taking.

Subjects in Group 4--study condition-no test, test condition-multiple-choice and Group 5--study condition-no test, test condition-short answer went with another experimenter to a vacant classroom where it was explained to them that one of their functions in the evaluation project previously discussed was to respond to the test their classmates were taking. All agreed to continue to participate and were administered, according to their respective group, either a multiple-choice or a short answer test.

Meanwhile, after explaining that the coordinator of the educational psychology classes was working with the students who had left the room, the instructor in the regular classroom routinely administered from one common stack either a multiple-choice or a short answer test respectively to students in Group 1--told test-multiple-choice and Group 2--told test-short answer. At the beginning of the next class meeting, students in Groups 1 and 2 were informed, as their classmates had been earlier, that their class was participating in an evaluation of instructional procedures which would be explained further at a later date.

Six weeks later, during which time the regularly scheduled classroom activities ensued, an experimenter came back to each class on an unannounced basis and administered both experimental tests to all students. They were told that their performance on these tests was the concluding part of the evaluation they had earlier been asked to participate in. Each person was strongly encouraged to do his or her best on both tests with an added incentive for conscientious effort being an "A" instead of the unit test scores.

Data Preparation and Analysis. In order to guard against bias in the scoring of the test, all identifying information was concealed and each test was assigned an identifying number. The multiple-choice tests were then scored using an objective scoring key. A detailed scoring key was prepared and used to score the short answer tests. In order to have uniformity throughout, one experimenter with the previously demonstrated ability to reliably score short answer tests ($r \geq .98$) scored all tests.

Resulting scores from the retention tests were analyzed using a 2 X 3 X 2 X 2 factorial analysis of variance with repeated measures on

the last two factors. The two between factors were study condition (test, no test) and test treatment condition (multiple-choice, short answer, no test). The within factors were item type on criterion measure (multiple-choice, short answer) and item complexity within the criterion test (knowledge, concept). For all significant interactions and for appropriate main effects Tuckey's HSD test was used for making pairwise comparisons of the means. All effects reported as significant were at the .05 level or less.

Results

The main effect for study condition was significant, F (1, 84) = 11.65, $p < .001$. Subjects in the study condition-test group ($\bar{X}$ = 8.79) scored higher than subjects in the study condition-no test group ($\bar{X}$ = 7.53). (Note: All means reported are an average of the means for levels of factors involved.) The main effect for the test treatment condition was significant, F (2, 84) = 4.95, $p < .01$. Although differences did exist among the multiple-choice treatment group ($\bar{X}$ = 8.825), the short-answer treatment group ($\bar{X}$ = 8.242), and the no test treatment group ($\bar{X}$ = 7.048), these differences were not explored due to the significant interaction between study condition and test treatment condition, F (2, 84) = 6.50, $p < .01$. Results of Tuckey's test revealed that study condition test subjects who took either the multiple-choice test ($\bar{X}$ = 9.58) or the short-answer test ($\bar{X}$ = 9.62) scored higher than subjects who took no test ($\bar{X}$ = 7.17), but the means for the multiple-choice test group and the short-answer test group did not significantly differ. Within the study condition-no test, however, means for subjects in the multiple-choice ($\bar{X}$ = 8.07), short-answer ($\bar{X}$ = 6.27) and no test ($\bar{X}$ = 7.65) treatment groups did not differ significantly.

Although there was a significant main effect for the first within group factor, criterion item type, F (1, 84) = 328.75, $p < .01$, knowing that students score higher on multiple-choice items than on short-answer items contributes little or no valuable information. However, the significant interaction between the within group factor item type (multiple-choice vs. short answer) and the between group factor test treatment condition (multiple-choice, short answer, and no test) is of importance, F (2, 84) = 3.38, $p < .05$. Again using Tuckey's HSD test, on the multiple-choice dependent measure subjects who took the multiple-choice test initially as a treatment (multiple-choice treatment group) ($\overline{X}$ = 11.08) scored higher than subjects who took the short-answer test as a treatment (short-answer test treatment group) ($\overline{X}$ = 9.85) and the subjects who received no test at the initial testing time (no test treatment group) ($\overline{X}$ = 9.58). The short-answer test treatment group failed to differ from the no test treatment group. Using the response to the short-answer questions as the dependent measure, a different pattern of subject responses was found. Subjects in the no test treatment group ($\overline{X}$ = 5.23) scored lower than subjects in the multiple-choice treatment group ($\overline{X}$ = 6.57) and subjects in the short-answer treatment group ($\overline{X}$ = 6.63). Subjects in the latter two groups failed to differ.

Criterion item type failed to interact with the between group factor of study condition.

With the final within group factor, item complexity, there was a significant main effect, F (1, 84) = 11.07, $p < .01$. Item complexity did not interact with either of the between group factors, test treatment condition and study condition. There was a significant interaction between the two within group factors, item complexity and item type, F (1, 84) = 59.45, $p < .001$.

However, finding that students scored higher on conceptual questions cast in a multiple-choice format but lower on conceptual items cast in a short-answer type item format was of minimum value to the study.

None of the three- and four-way interactions were significant.

Conclusions

The most profound conclusion supported by the results of this study is that students should both study for and take a test to maximize retention. Such a conclusion is diametrically opposed to observations such as those by John Holt (1964): "the test-examination-marks business is a gigantic racket, the purpose of which is to enable students, teachers, and schools to take part in a joint pretense" (p. 135). It certainly is not in line with recommendations such as those of Postman and Weingartner (1969) which call for a moratorium on the use of textbooks, elimination of all courses and all course requirements, and the abolition of tests and grades.

A secondary conclusion supported in this study is that, even though testing seems to influence retention, this influence is in part a joint function not only of the type of test initially administered but also the type of test used as a criterion measure.

Summary References

Harnquist, K. Enduring effects of schooling--A neglected area in educational research. Educational Researcher, 1977, 6, 5-11.

Holt, J. How children fail. New York: Pitman, 1964.

Postman, N., & Weingartner, C. Teaching as a subversive activity. New York: Delacorte Press, 1969.

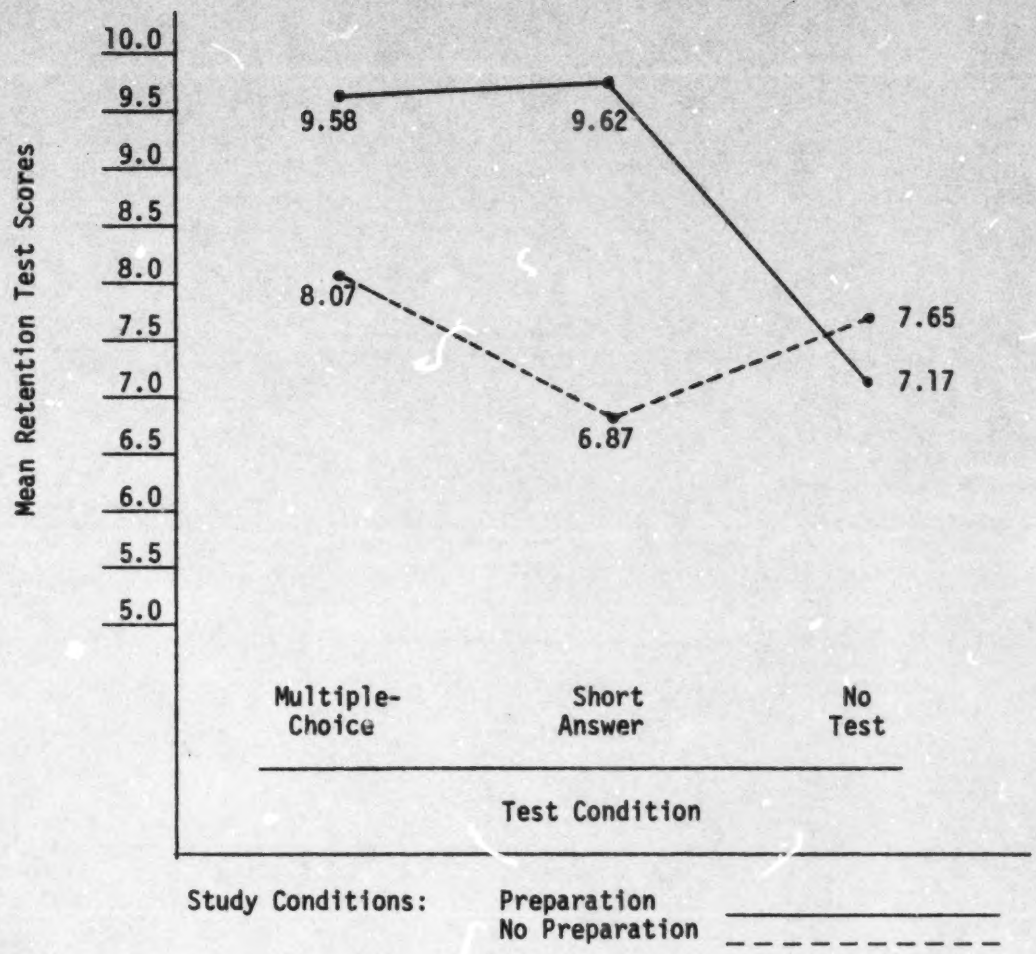

Figure 1. Interaction of Test Condition and Study Condition.

TABLE 1

ANOVA OF STUDY CONDITION, TEST CONDITION, ITEM TYPE, AND

ITEM COMPLEXITY FOR RETENTION TEST

Source	Sum of Squares	Degrees of Freedom	F Value
Study Condition	143.14	1	11.65***
Test Condition	121.67	2	4.95**
Study Condition X Test Condition	159.75	2	6.50**
Between Group Factor	1031.66	84	
Item Type	1460.05	1	328.75***
Item Type X Study Condition	7.80	1	1.76
Item Type X Test Condition	30.02	2	3.38*
Item Type X Study Condition X Test Condition	12.29	2	1.38
Within Group Error (1)	373.06	84	
Item Complexity	40.67	1	11.07***
Item Complexity X Study Condition	0.62	1	0.17
Item Complexity X Test Condition	2.02	2	0.28
Item Complexity X Study Condition X Test Condition	13.40	2	1.82
Within Group Error (2)	308.53	84	
Item Type X Item Complexity	95.07	1	59.45***
Item Type X Item Complexity X Study Condition	1.00	1	0.63
Item Type X Item Complexity X Test Condition	5.76	2	1.80
Item Type X Item Complexity X Study Condition X Test Condition	1.09	2	0.34
Within Group Error (3)	134.33	84	

***$p < .001$ **$p < .01$ *$p < .05$

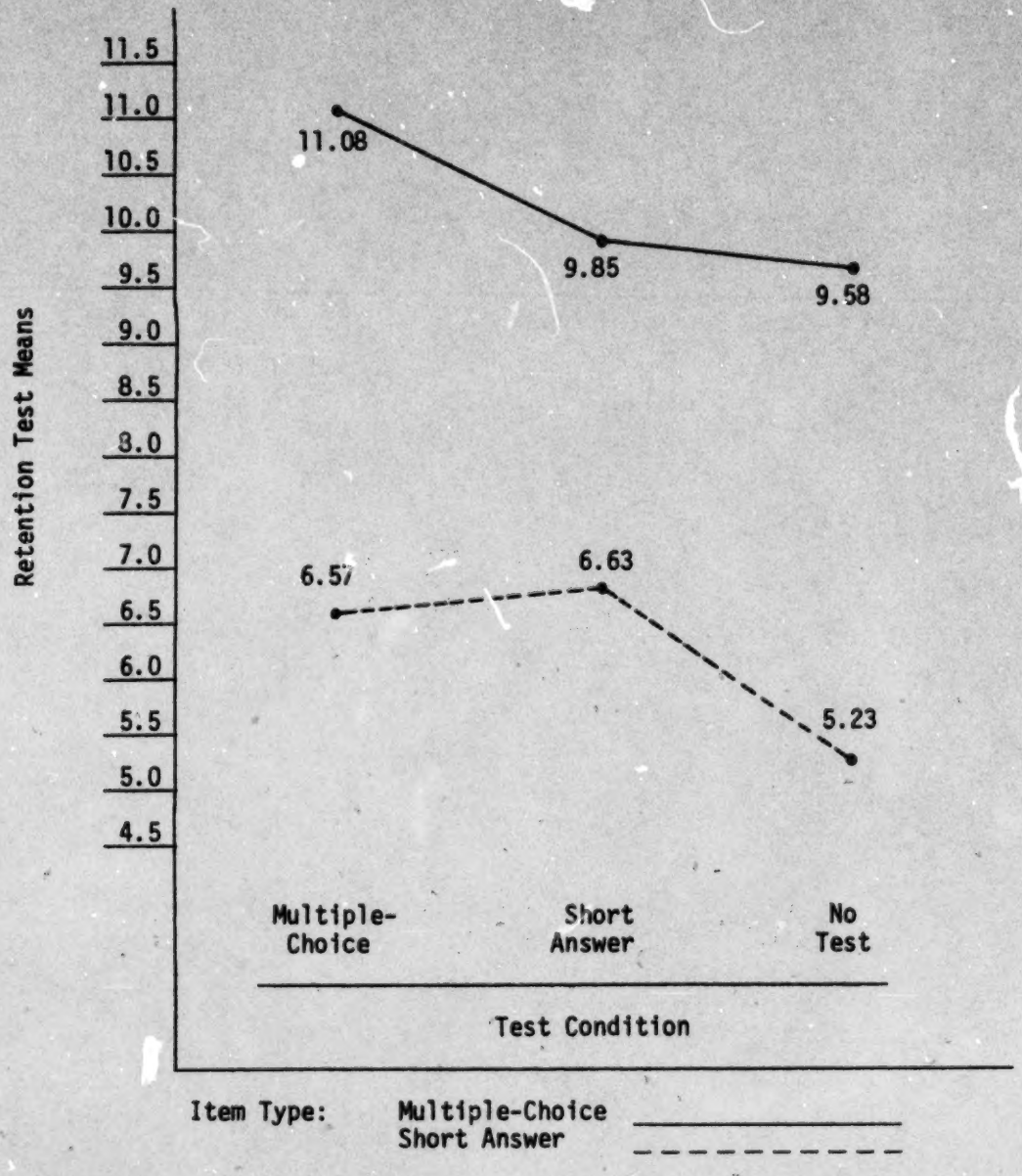

Figure 2. Interaction of Item Type and Test Treatment Condition.

TABLE 2

RETENTION TEST MEANS AND STANDARD DEVIATIONS FOR TREATMENT GROUPS

Retention Test	Prepared Multiple-Choice Mean	SD	Prepared Short Answer Mean	SD	Prepared No Test Mean	SD	Not Prepared Multiple-Choice Mean	SD	Not Prepared Short Answer Mean	SD	Not Prepared No Test Mean	SD
Multiple-Choice Knowledge	10.87	1.81	9.60	2.23	8.67	2.89	9.60	1.72	8.27	2.63	8.93	2.09
Multiple-Choice Concept	12.47	1.68	12.13	1.96	10.20	2.37	11.40	1.76	9.40	2.16	10.53	2.45
Short Answer Knowledge	7.80	2.60	8.47	1.77	5.07	1.83	5.87	2.26	5.53	1.92	5.20	2.54
Short Answer Concept	7.20	3.12	8.27	2.79	4.73	2.66	5.40	2.69	4.27	1.16	5.93	3.65